UNIVERSITY OF BIRMINGHAM
University of Birmingham Research at Birmingham

The colonoscopy Direct Observation of Procedural Skills assessment tool for evaluating competency development during training

Siau, Keith; Crossley , James; Dunckley, Paul; Johnson, Gavin; Feeney , Mark ; Iacucci, Marietta; Anderson , John T

DOI:
10.14309/ajg.0000000000000426

License:
None: All rights reserved

Document Version
Peer reviewed version

Citation for published version (Harvard):
Siau, K, Crossley , J, Dunckley, P, Johnson, G, Feeney , M, Iacucci, M & Anderson , JT 2019, 'The colonoscopy Direct Observation of Procedural Skills assessment tool for evaluating competency development during training', *The American Journal of Gastroenterology*. https://doi.org/10.14309/ajg.0000000000000426

Link to publication on Research at Birmingham portal

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

•Users may freely distribute the URL that is used to identify this publication.
•Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
•User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
•Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Download date: 06. May. 2023

THE COLONOSCOPY DOPS ASSESSMENT TOOL FOR EVALUATING COMPETENCY DEVELOPMENT DURING TRAINING

Keith **Siau**[1,2], James **Crossley**[3], Paul **Dunckley**[1,4], Gavin **Johnson**[1,5], Mark **Feeney**[1,6], Marietta **Iacucci**[7], John T **Anderson**[1,8]

On behalf of the Joint Advisory Group on Gastrointestinal Endoscopy (JAG)

Institutions:

[1] Joint Advisory Group on Gastrointestinal Endoscopy, Royal College of Physicians, London, UK
[2] Medical and Dental Sciences, University of Birmingham, Birmingham, UK
[3] Academic Unit of Medical Education, University of Sheffield, Sheffield, UK
[4] Department of Gastroenterology, Gloucestershire Hospitals NHSFT, Gloucester, UK
[5] Department of Gastroenterology, University College London Hospitals NHSFT, London, UK
[6] Department of Gastroenterology, Torbay and South Devon NHS Foundation Trust, Torquay, UK
[7] NIHR Biomedical Research Centre, University of Birmingham and University Hospitals NHS Foundation Trust Birmingham, UK
[8] Endoscopy, Gloucestershire Hospitals NHSFT, Cheltenham, UK

Corresponding author (and Guarantor): Dr Keith Siau; keithsiau@nhs.net

Short Title: Colonoscopy DOPS

Keywords:
Competence, Colonoscopy, DOPS, assessment, training

Contributions:
Study conception: KS, JC, PD, MF, JTA
DOPS development: PD, GJ, MF, JTA
Statistical analyses: KS, JC
Initial draft of manuscript: KS, JTA
Critical revisions: JC, PD, GJ, MF, MI, JTA

Potential Competing Interests: Keith Siau, Paul Dunckley, Gavin Johnson, Mark Feeney and John T Anderson are affiliated with the Joint Advisory Group on Gastrointestinal Endoscopy.

Word Count: 3308, Abstract: 259

ABSTRACT

Background: Formative colonoscopy direct observation of procedural skills (DOPS) assessments were updated in 2016 and incorporated into UK training but lack validity evidence. We aimed to appraise the validity of DOPS assessments, benchmark performance and evaluate competency development during training in diagnostic colonoscopy.

Methods: This prospective national study identified colonoscopy DOPS submitted over an 18-month period to the UK training e-portfolio. Generalisability analyses were conducted to evaluate internal structure validity and reliability. Benchmarking was performed using receiver operator characteristics (ROC) analyses. Learning curves for DOPS items and domains were studied and multivariable analyses performed to identify predictors of DOPS competency.

Results: Across 279 training units, 10749 DOPS submitted for 1199 trainees were analysed. The acceptable reliability threshold (G>0.70) was achieved with 3 assessors performing 2 DOPS each. DOPS competency rates correlated with the unassisted caecal intubation rate (rho 0.404, P<0.001). Demonstrating competency in 90% of assessed items provided optimal sensitivity (90.2%) and specificity (87.2%) for benchmarking overall DOPS competence. This threshold was attained in the following order: 'pre-procedure' (50-99 procedures), 'endoscopic non-technical skills' and 'post-procedure' (150-199), 'management' (200-249) and 'procedure' (250-299) domain. At item-level, competency in 'proactive problem solving' (rho 0.787) and 'loop management' (rho 0.780) correlated strongest with the overall DOPS rating (P<0.001) and were the last to develop. Lifetime procedure count, DOPS count, trainer specialty, easier case difficulty and higher caecal intubation rate were significant multivariable predictors of DOPS competence.

Conclusion: This study establishes milestones for competency acquisition during colonoscopy training and provides novel validity and reliability evidence to support colonoscopy DOPS as a competency assessment tool.

WHAT IS NEW HERE

What is known

- Standardised, validated, competency-assessment tools in colonoscopy training are lacking.

What is new here

- The DOPS can be used to measure development of individual competencies.
- Competency in DOPS mirrors the unassisted caecal intubation rate.
- DOPS is a valid and reliable tool which can be used to support colonoscopy training.

INTRODUCTION

Globally, renewed commitments towards the quality assurance of colonoscopy has led to a modern-day shift in the paradigm of colonoscopy training. The era of competency-based medical education has drawn competency evaluation away from merit based on numbers and towards continuous and objective evaluation of competence using standardised assessment tools.[1,2] There is an increasing expectation for training programmes to set competency benchmarks, facilitate objective and standardised assessment of competence, and to monitor and ensure competency development during training.[3] The implementation of objective, valid and reliable competency-assessment tools are necessary to support the delivery of this concept.[4]

The UK is one of the first countries to implement standardised assessment in colonoscopy. The Joint Advisory Group on Gastrointestinal Endoscopy (JAG),[5] the national quality assurance body for endoscopic procedures, supported the development of direct observation of procedural skills (DOPS) as an assessment tool. The first colonoscopy DOPS was developed in 2004 by a multidisciplinary panel of experts using a hierarchical task analysis format.[6] Over the years, the colonoscopy DOPS assessment form has undergone several iterations in response to feedback. In 2016, all DOPS assessment forms were modified to include: 1) revisions to the items and descriptors, 2) change in the scoring from a competency-based to supervision-based scale, and 3) the addition of a generic endoscopic non-technical skills assessment domain, which have improved the validity of competency assessments.[7] The current formative DOPS for colonoscopy (**Supplementary File 1**) contains 24 items, split into 5 domains, and an overall assessor rating.[8] The DOPS is accompanied by item descriptors (**Supplementary File 2**) which serve to facilitate objectivity of assessment.

Formative DOPS are completed by trainers after directly observing the performance of a trainee. DOPS are entered onto a national web-based system, implemented in 2009 for trainees to record endoscopy experience and assessment data - the JAG Electronic Training System (JETS) e-portfolio.[9]

Formative DOPS are defined by their purpose, i.e. assessments which highlight specific areas for development and monitor progress longitudinally. For trainees in the later-stages of training, formative assessments may also be used to gauge readiness for formal summative assessment, a gateway step for independent practice. The JAG criteria for provisional colonoscopy certification and suitability for independent colonoscopy practice comprise: lifetime colonoscopy count ≥200, unassisted caecal intubation rate ≥90% over the preceding 3 months, attendance of the JAG basic skills course in colonoscopy, and competency in the preceding 5 formative DOPS, as determined by each DOPS scoring ≥90% competency in assessed items.[10,11] Validity and reliability data for the previous iteration of colonoscopy DOPS had been reported,[6] but in the context of summative assessments for Bowel Cancer Screening accreditation (asymptomatic guaiac-positive patients). The study did not report on the benchmarking of outcomes and involved assessors who received five hours of dedicated DOPS training, which would be unfeasible to implement nationally. Thus, a study involving the updated DOPS in the formative (in-training) environment remains warranted. Given that each DOPS assesses up to 24 individual competencies, analyses of DOPS from a national training cohort can provide the granularity to study the development of specific technical, cognitive and generic non-technical skills during colonoscopy training. Such data may enable training programmes to benchmark performance and determine competency milestones.

In this national study involving DOPS assessments of colonoscopy trainees, we aimed to: i) assess the validity and reliability of formative DOPS, ii) analyse DOPS data to benchmark competence and evaluate competence development during training, and iii) identify independent predictors of DOPS competence.

METHODS

Study Design

This prospective, observational UK-wide study identified updated formative colonoscopy DOPS submitted onto the JETS e-portfolio over the 18-month period between July 2016-January 2018. Under JAG recommendations, the trainer's decision to perform a DOPS assessment is made prior to commencing a procedure and without the knowledge of the trainee, in order to minimise case-selection bias. Summative colonoscopy DOPS performed over this period were separately appraised.[12]

Factors Studied

For each DOPS completed, the following data were systematically collated: the individual item scores, case difficulty and overall DOPS rating awarded by the assessor/trainer. Other covariates studied included: the trainee and assessor identifier, trainee grade, lifetime colonoscopy count immediately preceding the DOPS assessment, lifetime sigmoidoscopy count, gastroscopy certification status, lifetime DOPS count and the unassisted caecal intubation rate calculated over the preceding 30 procedures.[10] Lifetime procedure counts were derived from trainee-entered procedures logged onto the JETS e-portfolio.

In the UK, the vast majority of all colonoscopy procedure are performed by three main specialties: gastroenterologists, GI surgeons and non-medical (nurse) endoscopists.[13] Gastroenterology specialty training currently lasts five years and runs concurrently with training in acute general medicine and on-call commitments. In surgery, lower gastroenterology training is a sub-speciality block within a longer more general surgical training programme. In gastroenterology, all training is usually associated with some commitment to general medical support throughout their training. Speciality training grades range from years 3 to 7 (ST3-7) and training in GI surgery lasts six years (ST3-8). It is possible to take time out at any stage of training for dedicated research, i.e. clinical research fellow role. Certification in colonoscopy is not a requirement for completion of specialist training and may

be continued upon completion as a consultant. Thus, for the purpose of analysis, ST3-5 grades were classified as junior, ST6+ as senior, and clinical research fellow treated as a separate category.

Outcomes

The primary outcome studied was the overall DOPS rating, which was awarded independent of the DOPS items scores and rated on a 4-point ordinal scale (**Supplementary File 3**).[8] These were converted to a 4 point numerical scale to facilitate analyses: Score 1: requiring maximal supervision; Score 2: significant supervision; Score 3: minimal supervision; Score 4: competent without supervision. Endoscopic competence was defined as the ability to perform a procedure in an effective, safe and timely manner. Items could be rated "not applicable" (N/A) if assessment was possible. N/A scores were excluded from item and domain-level analyses. The percentage of items scoring competent (Score 4) for each DOPS and domain (pre-procedural, technical, post-procedural and ENTS) were adopted as the secondary outcome.

Statistical Analyses

Item-total correlations

Item-total analyses can be used to indicate internal structure validity by indicating differences in magnitude of correlation between each DOPS item and the overall DOPS score. Analyses were performed using Spearman's rank tests, with rho coefficients ≥0.70 regarded as a strong positive correlation.

Generalisability Theory

The reliability of DOPS assessments was evaluated using generalisability theory, a statistical framework which applies variance component analysis to determine the influence of independent variables on assessment outcomes, i.e. overall DOPS score.[14,15] For this assessment design, the key variance components are: trainee ability (across all assessors and cases: $V_{trainee}$), assessor stringency (across all trainees and cases: $V_{assessor}$), assessor subjectivity attributable to the trainee ($V_{assessor*trainee}$), and residual variance (V_{error}), most of which will be trainee case-to-case variation. This enables generalisability coefficients (G) to be calculated as a function of the number of cases and assessors. The G-coefficient is calculated using the same general equation as a reliability coefficient (G= trainee variance / [trainee variance + error variance]), with values ≥0.70 considered of acceptable reliability for in-training assessments.

Benchmarking Competency

Competency benchmarks were derived from receiver operating characteristics (ROC) curve analyses to determine the optimal percentage of competent items per DOPS in relation to overall DOPS competency. The value providing optimal sensitivity and specificity was calculated using Youden's index (sensitivity + specificity - 1).[16]

Competency Development

To estimate competency development (learning curves), the percentage of competent scores were calculated for each item, domain and global measure (overall assessor rating and percentage of competent items per DOPS), and stratified by lifetime procedure count. The lifetime procedure counts required to reach the competency benchmark score was used to estimate when trainees develop specific competencies as measured within DOPS.

Predictors of DOPS Competence

Multivariable analyses using a binary logistic regression approach were then performed to identify independent factors associated with DOPS competence (overall DOPS outcome of 4). Generalised estimating equations (GEE) with an autoregressive (AR1) structure were used to account for the non-independence of trainee procedures. DOPS with missing data and those performed by NME trainees were excluded from analyses, as unlike other specialties, NME trainees are usually trained to deliver independent sigmoidoscopy practice, e.g. in Bowel Scope screening, before training in colonoscopy, and such data may not be captured on the JETS e-portfolio.

Statistical analyses were performed in SPSS (v24, Arkmont, NY: IBM Corp), with $p<0.05$ indicative of significance throughout.

RESULTS

Participants

During the study period a total of 10749 DOPS were included for analysis. DOPS were submitted by 1399 trainers (median 4 DOPS per trainer; IQR 2-9) for 1199 trainees (median 5 DOPS per trainee; IQR 2-13), within 279 UK endoscopy training units. Trainee characteristics are summarised in **Table 1**. Those within the NME specialty had performed more flexible sigmoidoscopy procedures ($P<0.001$) prior to DOPS assessments (median 92, IQR: 20-265) than gastroenterology (median 44, IQR 44-133) and GI surgery trainees (median 51, IQR 17-104).

Item-Total Correlations

Item-total correlations were conducted to assess the relationship between the performance of each DOPS item and the overall competency rating. This showed positive and significant correlations (P<0.001) for all items (**Table 2**). Item-total correlations were weakest for items located within the 'pre-procedure' domain, and strongest for those within the 'procedure' domain, particularly for 'proactive problem solving' (rho 0.787), 'loop management' (rho 0.780), 'pace and progress' (rho 0.734) and 'tip control' (rho 0.719).

Sources of Variance

Variance component analysis yielded the following estimates of the effect of key variables on the overall DOPS assessor rating: trainee ability (38%), assessor stringency (17%), assessor subjectivity attributable to the trainee (12%) and residual variance (34%).

Reliability

Combining the variance estimates based on generalisability theory, the reliability of formative DOPS was modelled from different combinations of trainers and observations (**Table 3**). 6 observations (≥2 DOPS each from 3 different assessors) was sufficient to achieve the reliability threshold of 0.70 for in-training assessment.

Competency Thresholds

ROC analysis (**Figure 3**) showed that overall competency for an assessed procedure could be predicted by each component assessed item within DOPS, i.e. based on the percentage of items rated as competent within each DOPS (area under ROC: 0.956, P<0.001). Attainment of competency in 90% of assessed DOPS items yielded optimal sensitivity (90.2%) and specificity (87.2%) for predicting overall

competence, with corresponding false positive and false negative rates of 12.8% and 9.8% respectively. The DOPS competency rate of 90% was therefore set as the benchmark for delineating competence.

Competency Development during Colonoscopy Training

Across the cohort, DOPS performance, as measured by the mean percentage of competent scores for each item (**Table 4**) and domain (**Figure 1**), was presented by strata of lifetime procedure count. Considering all DOPS items, the 90% competency threshold was achieved after 200-249 procedures. Trainees first developed competencies in the "pre-procedure" domain, in the item order of "consent", "preparation" and "equipment check" (50-99 procedures), before other "pre-procedure" items (100-149 procedures). Competency development profiles for "post-procedure" and "ENTS" domains were similar and developed after 150-199 procedures. This was followed by competence in the "management" domain and in the "leadership" and "judgement and decision making" items within the ENTS domain (200-249 procedures). Finally, proficiency in the "procedural" domain were attained after 250-299 procedures. At item-level, competencies in "air management", "visualisation", "patient comfort" and "scope handling" were achieved by 200-249 procedures. After 300 procedures, the competency threshold was reached for "tip control" (90.8%) and "pace and progress" but not for "proactive problem solving" (88.1%) and "loop management" (86.1%). Competency development profiles did not vary by case difficulty (**Supplementary File 4**), although more complicated cases were associated with lower rates of competent item scores within the "procedure" domain. Comparisons between trainee specialties for each block of lifetime procedure count (**Supplementary Files 5-7**) suggest that surgical trainees are more likely to be awarded DOPS competency than gastroenterology counterparts.

Trainees achieved the median unassisted CIR target of 90% after 200-249 procedures (**Figure 2**); moderately positive correlation was observed between CIR and DOPS performance (rho 0.404, P<0.001).

DOPS Competency

On multivariable GEE analysis (**Table 5**), lifetime procedural count (P<0.001), case difficulty (P<0.001), lifetime colonoscopy count (P<0.001), lifetime colonoscopy DOPS count (P=0.002), and higher unassisted caecal intubation rate (P<0.001) were independently associated with DOPS competence. There was a direct relationship between assessor specialty and DOPS competence (P=0.002). After accounting for these factors, trainee specialty (P=0.665), trainee seniority (P=0.220), gastroscopy certification (P=0.161) and lifetime flexible sigmoidoscopy count (P=0.681) were not significantly associated with DOPS competence.

Over the study period, 184 trainees (gastroenterologist: 78; gastrointestinal surgeon: 68; non-medical endoscopist:35) applied for summative DOPS assessments for provisional colonoscopy certification after meeting eligibility criteria. The median lifetime colonoscopy count prior to summative assessment was 236 (IQR 211-285), with no significant differences between the three specialties (P=0.210). In total, 183 of the 184 trainees (99.5%) were deemed competent over the 4 summative DOPS assessments and were awarded certification for independent colonoscopy practice.

DISCUSSION

In this prospective nationwide study of colonoscopy DOPS assessments undertaken within 279 UK training centres, we present validity and reliability data to support the use of colonoscopy DOPS as an in-training assessment tool. Based on Messick's contemporary framework for evaluating validity in

assessment tools,[17,18] internal structure evidence (associations between test measures and reliability) was supported through item-total correlations and generalisability theory analyses, while learning curves and the observation that DOPS performance closely mirrors the unassisted CIR provides discriminative validity (relations to other variables). ROC analyses showed that all items measured within each DOPS could be accommodated to predict overall competence (AUROC=0.956); the evaluation of a competency threshold and its false positive and false negative rates attests to consequential validity (value implications of interpreting test scores). Content validity (relevance) may be inferred from the multidisciplinary nature of DOPS construction and the improvement in the DOPS rating scale,[7] whereas response process validity (relationship between intended construct and thought processes of assessors) stems from high trainer satisfaction reported with the previous iteration of colonoscopy DOPS.[6]

Over the last decade, a number of competency assessment tools have been developed and validated to support colonoscopy training.[2,3,19] The Assessment of Competency in Endoscopy (ACE) tool,[20] which succeeded the Mayo Colonoscopy Skills Assessment Tool (MCSAT), is currently used in North America.[21] ACE measures 14 competency items and awards overall scores for technical and cognitive skills on a performance-based scale ranging from 1 (novice) to 4 (highly-skilled). Following the analysis of 1061 ACE assessments for 93 fellows, an overall score of 3.5 was reported as the optimal competency threshold; this was attained for cognitive endpoints before technical skills, with competency achieved in nearly all items after 250 procedures.[22] The Gastrointestinal Endoscopy Competency Assessment Tool (GIECAT),[23] and more recently SAFE-T,[24] have also been introduced and supported by validity and reliability data, but not have been incorporated into national endoscopy training systems as with the DOPS.

The use of purpose-specific and validated formative assessment tools not only allows trainees and trainers to facilitate targeted feedback, but also enables training programmes to set performance

benchmarks and profile competency development. The 24 items measured within each colonoscopy DOPS provides the granularity to assess trainee progression in specific competency items. Data on day-to-day, real-world evaluations of competency development using formative assessment tools are lacking, especially those referencing the development of ENTS during colonoscopy training. In this study, the 90% competency threshold was attained in order of: 'pre-procedure' (50-99 procedures), 'ENTS' and 'post-procedure' (150-199), 'management' (200-249) and 'procedure' (250-299) domain. Such data appear to contrast that for gastroscopy where ENTS competencies developed after technical competencies.[25] This finding could reflect the demands of colonoscopy as a more technically-demanding procedure. As colonoscopy training may be preceded by training in gastroscopy, it is possible that some of the generic competencies, e.g. ENTS, may be transferrable between endoscopic modalities. At item-level, competencies in 'proactive problem solving' (rho 0.787) and 'loop management' (rho 0.780) correlated strongest with the overall DOPS rating (P<0.001) and were the last to develop. This approach identifies specific skills which could be considered for prioritisation. It is possible that, with future evolution, ENTS and other more recently added items will be further refined, which may influence the observed rate at which the inter-related endoscopy skills develop in parallel.

Competency development during colonoscopy training has been profiled using completion-based endpoints such as CIR, although its definition and calculation method may vary internationally.[26] Previous data from the UK JETS e-portfolio suggest that 233 lifetime procedures were required to achieve 90%+ unassisted CIR,[27] with other studies citing between 141-300 procedures to achieve this target.[28,29] Consistent with data from ACE assessments reported by Sedlack et al,[22] our trainees developed cognitive competence on procedure "indications" after 100 procedures. The ACE cohort reached competency thresholds for motor and cognitive endpoints after 250 procedures. For DOPS, the 200-249 lifetime procedure count marked the point where trainees began to achieve consistent unassisted CIRs of 90%+. At present, JAG mandates a minimum lifetime procedure count of 200 in

order to trigger summative assessment.[10] Despite this, previous JETS analyses found that trainees applied for provisional colonoscopy certification after a median of 269 (IQR 226 – 342) lifetime procedures.[13] Furthermore, it is evident that trainees certified closer to the 200 threshold are more likely to exhibit a drop in performance during newly-independent practice.[11] As such, mandating a minimum of 250 lifetime procedures may be suggested as a more appropriate criterion for triggering summative assessment. For items of "proactive problem solving" and "loop resolution", the 90% threshold was not accomplished even after 300 procedures, suggesting that these skills would continue to develop during the independent (post-credentialing) period. This may also explain why in more difficult cases, lower rates of competent item scores are seen within the "procedure" domain.

Our study had several limitations. First, the colonoscopy DOPS is primarily centred on assessing a trainee's ability to perform high-quality diagnostic colonoscopy. Competencies relevant to polypectomy are measured on a separate instrument: the direct observation of polypectomy skills (DOPyS).[30] Second, this was a real-world observational study of unselected cases involving all trainees from all centres within the UK. Colonoscopy in the UK is typically performed either under conscious sedation (with midazolam +/- intravenous opioid) or unsedated, which may be more challenging than training in healthcare settings where propofol is routinely administered. Third, competency development was assessed at the level of each DOPS assessment rather than each trainee. As the frequency and intervals of DOPS assessments varied amongst trainees, it was not appropriate to subject these to trainee-level analyses, except for the GEE regression models. Fourth, the DOPS does not collect procedure-specific data, e.g. sedation use, indication, diagnoses, complication rates etc. These are currently collected on the JETS e-portfolio but are not cross-linked with DOPS. Efforts are underway to integrate DOPS outcomes with procedural-level data. Lifetime colonoscopy counts were deduced from trainee-entered procedure entries on the JETS e-portfolio, which may risk selection bias and underestimate time-to-competency data. This is now being being addressed in the UK National Endoscopy Database project where all procedures are centrally recorded.[31] Finally, subgroup analyses

of DOPS competency rates were presented for gastroenterology and surgical specialties (**Supplementary File 5-7**). This data should be interpreted with caution. Although there may have been a trend for earlier competency development in surgical trainees, this difference did not remain after multivariable GEE analysis (**Table 5**). This analysis also uncovered assessor variation, whereby assessors from the GI surgical and non-medical (nurse) endoscopist specialties were more likely to rate DOPS as being competent than those from the gastroenterology specialty. This raises the possibility of assessor bias across trainer specialties. It is unclear whether this could reflect potential differences in caseload, efficacy of training, or variation in assessor training in the use of DOPS. We suspect that the stringency of gastroenterology assessors may arise from greater familiarity with DOPS, in particular from attendance of Train-the-Trainers courses and Bowel Cancer Screening accreditation where there is formal DOPS training. In these groups, there are proportionately far greater numbers of gastroenterologists who train when compared to surgeons who train. Further studies are needed to explore this further, in particular trainers' perceptions on the use of DOPS, and to determine whether trainer interventions are necessary to improve the consistency of DOPS scoring.

Few competency assessment tools have undergone such rigorous evaluation to identify potential heterogeneity on their usage. Multivariable regression models showed that, in addition to lifetime procedure count and assessor specialty, other variables associated with DOPS competence included DOPS count and easier case difficulty. Invariably, competency requires the demonstration of technical and non-technical proficiencies across different cases contexts and procedural difficulties.[3] Mandating higher combinations of trainers and assessments can improve the reliability of DOPS, with sufficient reliability (G>0.70) achievable with ≥3 assessors observing ≥2 DOPS each. Moreover, engagement with the formative assessment process can promote feedback and reflective practice, which may underpin the association between lifetime DOPS count and DOPS competency. With the current spotlight fixed on high-quality colonoscopy in the remit of competency-based curricula, training programmes are duty bound to ensure that competence is achievable, and that competency

elicited during the assessment process will effectively translate to competent clinical practice. The delivery of this concept is supported by the observation that 99.5% of trainees who performed competently within their last 5 formative DOPS, in conjunction with meeting minimum performance indicators, also passed their summative assessments required for certification. Optimising the validity and reliability of formative DOPS could potentially consolidate its role as an adjunctive training aid, and in experienced trainees, as a pseudo-summative competency safeguard to quality assure training in colonoscopy.

In summary, we provide validity and reliability evidence to support formative colonoscopy DOPS as an in-training assessment tool, and present data on competency development, as measured using DOPS.

REFERENCES

1. Nasca TJ, Philibert I, Brigham T, et al. The Next GME Accreditation System — Rationale and Benefits. 2012;366(11):1051-56.
2. Ekkelenkamp VE, Koch AD, de Man RA, et al. Training and competence assessment in GI endoscopy: a systematic review. Gut 2016;65(4):607-15.
3. Siau K, Hawkes ND, Dunckley P. Training in Endoscopy. Curr Treat Options Gastroenterol 2018;16(3):345-61.
4. Walsh CM. In-training gastrointestinal endoscopy competency assessment tools: Types of tools, validation and impact. Best Pract Res Clin Gastroenterol 2016;30(3):357-74.
5. Siau K, Green JT, Hawkes ND, et al. Impact of the Joint Advisory Group on Gastrointestinal Endoscopy (JAG) on endoscopy services in the UK and beyond. 2019;10(2):93-106.
6. Barton JR, Corbett S, van der Vleuten CP. The validity and reliability of a Direct Observation of Procedural Skills assessment tool: assessing colonoscopic skills of senior endoscopists. Gastrointest Endosc 2012;75(3):591-7.
7. Siau K, Dunckley P, Valori R, et al. Changes in scoring of Direct Observation of Procedural Skills (DOPS) forms and the impact on competence assessment. Endoscopy 2018;50(8):770-78.
8. Joint Advisory Group on Gastrointestinal Endoscopy. Formative DOPS: Colonoscopy and Flexible Sigmoidoscopy. 2016 [cited 20th April 2019]; Available from: https://www.thejag.org.uk/Downloads/DOPS%20forms%20(international%20and%20reference%20use%20only)/Formative%20DOPS_Colonoscopy%20and%20Flexible%20sigmoidoscopy.pdf.
9. Mehta T, Dowler K, McKaig BC, et al. Development and roll out of the JETS e-portfolio: a web based electronic portfolio for endoscopists. Frontline Gastroenterology 2010;2(1):35.
10. Joint Advisory Group on Gastrointestinal Endoscopy. JAG trainee certification process: Colonoscopy (provisional and full). 2017 [cited 20th April 2019]; Available from: https://www.thejag.org.uk/Downloads/JETS%20-%20certification%20for%20trainees/Colonoscopy%20application%20criteria%20and%20process.pdf.
11. Siau K, Hodson J, Valori RM, et al. Performance indicators in colonoscopy after certification for independent practice: outcomes and predictors of competence. Gastrointest Endosc 2019;89(3):482-92.e2.
12. Joint Advisory Group on Gastrointestinal Endoscopy. Summative DOPS: Colonoscopy and Flexible Sigmoidoscopy. 2016 [cited 20th April 2019]; Available from: https://www.thejag.org.uk/Downloads/DOPS%20forms%20(international%20and%20reference%20use%20only)/Summative%20DOPS_Colonoscopy%20and%20Flexible%20sigmoidoscopy.pdf.
13. Siau K, Anderson JT, Valori R, et al. Certification of UK gastrointestinal endoscopists and variations between trainee specialties: results from the JETS e-portfolio. Endosc Int Open 2019;7(4):E551-E60.
14. Crossley J, Davies H, Humphris G, et al. Generalisability: a key to unlock professional assessment. Med Educ 2002;36(10):972-8.
15. Webb NM, Shavelson RJ, Haertel EH. 4 Reliability Coefficients and Generalizability Theory. In: Rao CR, Sinharay S, eds. Handbook of Statistics: Elsevier 2006:81-124.
16. Ruopp MD, Perkins NJ, Whitcomb BW, et al. Youden Index and optimal cut-point estimated from observations affected by a lower limit of detection. Biom J 2008;50(3):419-30.
17. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American psychologist 1995;50(9):741.
18. Cook DA, Hatala R. Validation of educational assessments: a primer for simulation and beyond. Advances in Simulation 2016;1(1):31.
19. Leung FW. Assessment of trainees' performance in colonoscopy. Gastrointest Endosc 2018;87(1):270-71.
20. Fried GM, Marks JM, Mellinger JD, et al. ASGE's assessment of competency in endoscopy evaluation tools for colonoscopy and EGD. Gastrointest Endosc 2014;80(2):366-7.
21. Sedlack RE. The Mayo Colonoscopy Skills Assessment Tool: validation of a unique instrument to assess colonoscopy skills in trainees. Gastrointest Endosc 2010;72(6):1125-33, 33.e1-3.
22. Sedlack RE, Coyle WJ. Assessment of competency in endoscopy: establishing and validating generalizable competency benchmarks for colonoscopy. Gastrointest Endosc 2016;83(3):516-23.e1.
23. Walsh CM, Ling SC, Khanna N, et al. Gastrointestinal Endoscopy Competency Assessment Tool: reliability and validity evidence. Gastrointest Endosc 2015;81(6):1417-24.e2.

24. Kumar NL, Kugener G, Perencevich ML, et al. The SAFE-T assessment tool: derivation and validation of a web-based application for point-of-care evaluation of gastroenterology fellow performance in colonoscopy. Gastrointest Endosc 2018;87(1):262-69.
25. Siau K, Crossley J, Dunckley P, et al. Direct observation of procedural skills (DOPS) assessment in diagnostic gastroscopy: nationwide evidence of validity and competency development during training. Surg Endosc 2019 Published Online First: 2019/03/27. doi: 10.1007/s00464-019-06737-7.
26. Siau K, Hodson J, Ravindran S, et al. Variability in cecal intubation rate by calculation method: a call for standardization of key performance indicators in endoscopy. Gastrointest Endosc 2019;89(5):1026-36.e2.
27. Ward ST, Mohammed MA, Walt R, et al. An analysis of the learning curve to achieve competency at colonoscopy using the JETS database. Gut 2014;63(11):1746.
28. Sedlack RE. Training to competency in colonoscopy: assessing and defining competency standards. Gastrointestinal Endoscopy 2011;74(2):355-66.e2.
29. Cohen J, Cohen SA, Vora KC, et al. Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. Gastrointestinal Endoscopy 2006;64(3):361-68.
30. Joint Advisory Group on Gastrointestinal Endoscopy. Formative DOPyS: Colonoscopy and Flexible Sigmoidoscopy. 2016 [cited 21/11/17; Available from: https://www.thejag.org.uk/Downloads/DOPS%20Forms%20For%20International%20and%20reference%20use%20only/Formative%20DOPyS_Colonoscopy%20and%20Flexible%20sigmoidoscopy.pdf.
31. Lee TJW, Siau K, Esmaily S, et al. Development of a national automated endoscopy database: The United Kingdom National Endoscopy Database (NED). United European Gastroenterology Journal 2019;7(6):798-806.

FIGURE CAPTIONS

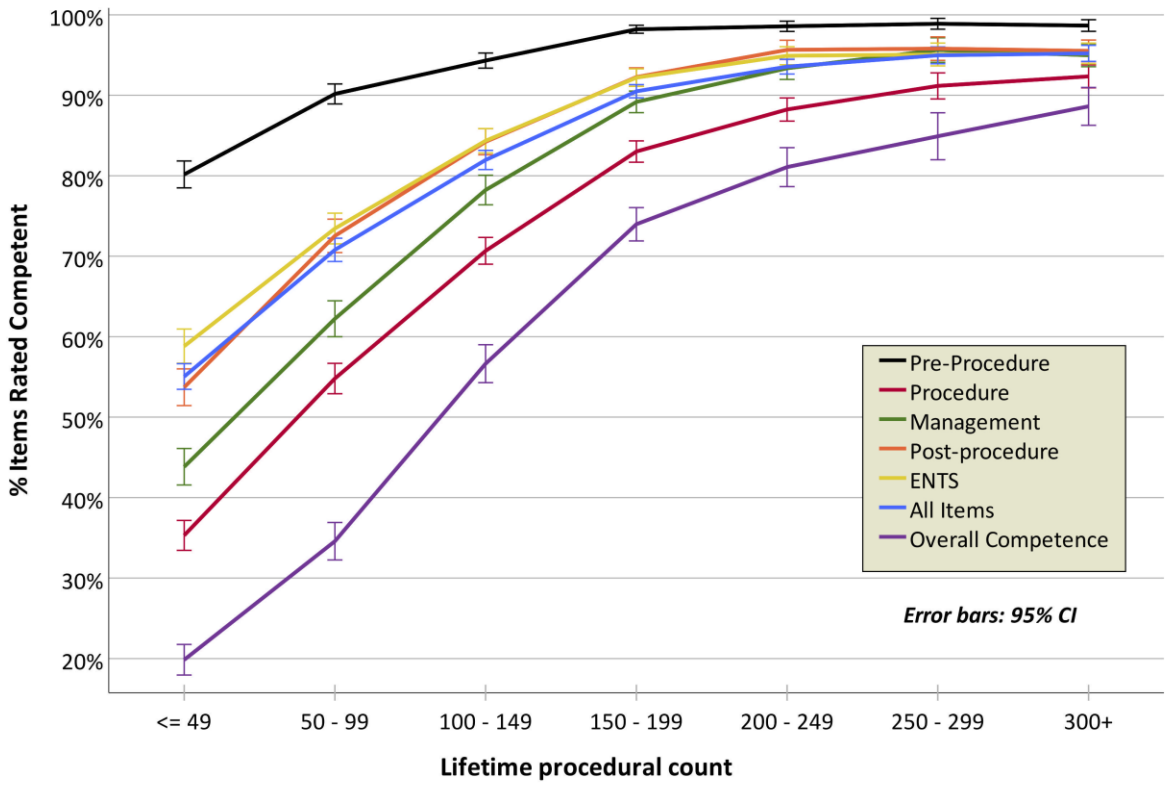


Figure 1: Learning curves in colonoscopy, as assessed by overall DOPS scores and average domain competency rates.

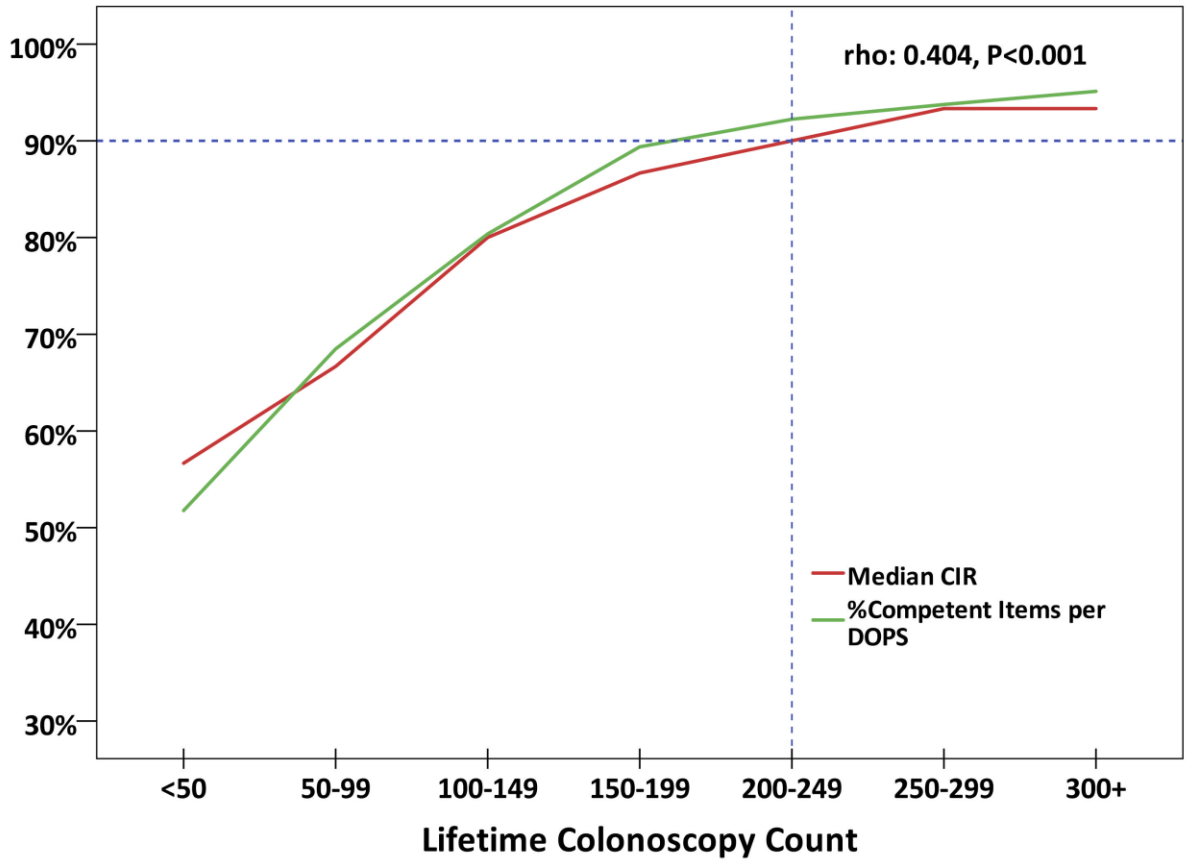
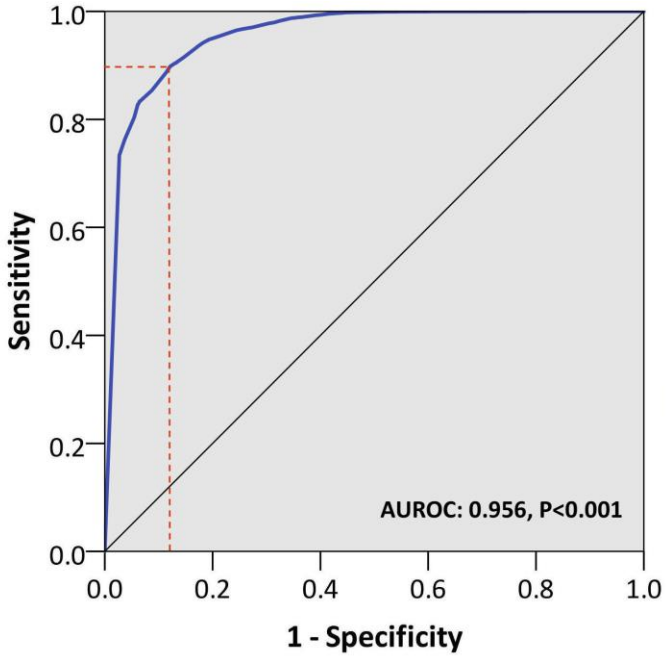


Figure 2: Correlation between median caecal intubation rate and percentage of competent items per DOPS, grouped by lifetime colonoscopy count.



Competence Threshold	Sensitivity	Specificity	Youden Index
80%	0.968	0.743	0.711
81%	0.965	0.755	0.720
82%	0.963	0.763	0.726
83%	0.953	0.792	0.745
84%	0.948	0.807	0.755
85%	0.946	0.809	0.755
86%	0.941	0.819	0.760
87%	0.935	0.828	0.763
88%	0.906	0.866	0.772
89%	0.905	0.867	0.772
90%	0.902	0.872	0.774
91%	0.887	0.886	0.773
92%	0.834	0.934	0.768
93%	0.834	0.934	0.768
94%	0.834	0.935	0.769
95%	0.829	0.937	0.766
96%	0.762	0.963	0.725
98%	0.734	0.973	0.707

Figure 3: Receiver operating characteristics curve (ROC) analysis demonstrating the relationship between DOPS competence threshold (percentage of assessed items on DOPS scoring competent) and overall DOPS competency rating. The competence threshold of 90% provided optimal sensitivity (90.2%) and specificity (87.2%) for competency determination. AUROC: area under receiver operating characteristics curve.

TABLES

	Number of Trainees (%)	Number of DOPS (%)
Lifetime Procedure Count*		
<50	316 (26.4%)	2299 (21.4%)
50-99	172 (14.3%)	1952 (18.2%)
100-149	122 (10.2%)	1996 (18.6%)
150-199	176 (14.7%)	1953 (18.2%)
200-249	152 (12.7%)	1115 (10.4%)
250-299	91 (7.6%)	659 (6.1%)
300+	170 (14.2%)	775 (7.2%)
Specialty		
Gastroenterology	524 (43.7%)	4969 (46.2%)
GI Surgeon	426 (35.5%)	3076 (28.6%)
Non-medical endoscopist	203 (16.9%)	2412 (22.4%)
Unknown/Other	46 (3.8%)	271 (2.5%)
Trainee Grade**		
ST3	61 (6.5%)	303 (3.8%)
ST4	135 (14.4%)	1268 (16.0%)
ST5	162 (17.2%)	1719 (21.7%)
ST6	144 (15.3%)	1209 (15.3%)
ST7	103 (11.0%)	760 (9.6%)
ST8	86 (9.1%)	591 (7.5%)
Clinical Research Fellow	52 (5.5%)	391 (4.9%)
Associate Specialist	110 (11.7%)	986 (12.4%)
Consultant	87 (9.3%)	698 (8.8%)

Table 1: Baseline characteristics of the colonoscopy training cohort. *Denotes the maximum lifetime procedure count preceding DOPS for each unique trainee. **For trainees with multiple training grades, the most recent allocation was used. ST: specialist trainee (ST3 usually refers to the first year of endoscopy training); GI: Gastrointestinal

DOPS Item	N	Spearman coefficient (rho)	P-value
Indication	10424	0.403	<0.001
Risk	10393	0.401	<0.001
Confirms consent	10342	0.336	<0.001
Preparation	10581	0.363	<0.001
Equipment check	10697	0.357	<0.001
Monitoring	10668	0.366	<0.001
Sedation	10075	0.409	<0.001
Scope handling	10748	0.692	<0.001
Tip control	10747	**0.719**	<0.001
Air management	10691	0.649	<0.001
Proactive problem solving	10587	**0.787**	<0.001
Loop management	10525	**0.780**	<0.001
Patient comfort	10719	0.646	<0.001
Pace and progress	10727	**0.734**	<0.001
Visualisation	10722	0.692	<0.001
Recognition	9801	0.642	<0.001
Management	9065	0.656	<0.001
Complications	3653	0.662	<0.001
Report writing	9032	0.533	<0.001
Management plan	9502	0.540	<0.001
Communication and teamwork	10682	0.507	<0.001
Situation awareness	10634	0.543	<0.001
Leadership	10361	0.570	<0.001
Judgement and decision making	10460	0.617	<0.001

Table 2: Correlation between colonoscopy DOPS items and overall competence rating (Item-global correlations) presented as Spearman's rho coefficients. *Strongly positive correlations (rho >0.70) indicated in bold.*

		Observations per Trainer							
		1	**2**	**3**	**4**	**5**	**10**	**15**	**20**
Trainers	**1**	0.38	0.45	0.48	0.50	0.51	0.54	0.55	0.55
	2	0.55	0.62	0.65	0.67	0.68	**0.70**	**0.71**	**0.71**
	3	0.64	**0.71**	**0.74**	**0.75**	**0.76**	**0.78**	**0.78**	**0.79**
	4	**0.71**	**0.77**	**0.79**	**0.80**	**0.81**	**0.82**	**0.83**	**0.83**
	5	**0.75**	**0.80**	**0.82**	**0.83**	**0.84**	**0.85**	**0.86**	**0.86**
	6	**0.78**	**0.93**	**0.85**	**0.86**	**0.88**	**0.88**	**0.88**	**0.88**

Table 3: Reliability estimates (G-coefficients) of formative colonoscopy DOPS based on 1-6 trainers each observing 1-20 assessments. *G-coefficients of 0.70+ based on assessor and assessment combinations (indicating sufficient reliability for in-training assessment) are shown in bold.*

	Lifetime Procedure Count						
	<50	**50-99**	**100-149**	**150-199**	**200-249**	**250-299**	**300+**
Indication	75.4% (73.6%-77.2%)	87.5% (86.0%-88.9%)	**93.6% (92.4%-94.6%)**	**98.0% (97.3%-98.5%)**	**97.5% (96.5%-98.3%)**	**98.9% (97.9%-99.5%)**	**98.8% (97.8%-99.4%)**
Risk	75.8% (74.0%-77.6%)	87.0% (89.9%-93.7%)	**93.7% (92.5%-94.7%)**	**97.9% (97.1%-98.4%)**	**97.9% (96.9%-98.6%)**	**99.1% (98.1%-99.6%)**	**98.3% (97.2%-99.0%)**
Confirms consent	82.4% (80.8%-84.0%)	**91.6% (90.3%-92.8%)**	**95.6% (94.6%-96.4%)**	**98.5% (97.9%-99.0%)**	**98.5% (97.7%-99.1%)**	**99.4% (98.5%-99.8%)**	**99.2% (98.4%-99.7%)**
Preparation	79.4% (77.7%-81.1%)	**90.5% (89.2%-91.8%)**	**94.4% (93.4%-95.4%)**	**98.2% (97.5%-98.7%)**	**98.3% (97.4%-98.9%)**	**99.1% (98.1%-99.6%)**	**99.2% (98.4%-99.7%)**
Equipment check	79.3% (77.6%-80.9%)	**90.2% (88.8%-91.5%)**	**94.3% (93.2%-95.3%)**	**97.7% (97.0%-98.3%)**	**97.9% (97.0%-98.6%)**	**98.2% (96.9%-99.0%)**	**98.8% (97.9%-99.4%)**
Monitoring	78.6% (76.9%-80.3%)	89.3% (87.9%-90.6%)	**93.9% (92.7%-94.9%)**	**98.1% (97.4%-98.6%)**	**98.4% (97.5%-99.0%)**	**98.8% (97.7%-99.4%)**	**99.1% (98.2%-99.6%)**
Sedation	70.5% (68.5%-72.4%)	85.7% (84.1%-87.3%)	**90.9% (89.5%-92.1%)**	**96.9% (96.0%-97.6%)**	**97.3% (96.2%-98.1%)**	**98.3% (97.0%-99.1%)**	**97.6% (96.3%-98.5%)**
Scope handling	35.5% (33.5%-37.4%)	58.8% (56.6%-61.0%)	74.0% (72.1%-75.9%)	85.6% (84.0%-87.1%)	**90.8% (89.0%-92.4%)**	**93.3% (91.2%-95.0%)**	**96.0% (94.4%-97.2%)**
Tip control	29.4% (27.6%-31.3%)	50.6% (48.4%-52.9%)	68.2% (66.2%-70.3%)	80.5% (78.7%-82.2%)	84.6% (82.4%-86.6%)	87.9% (85.2%-90.2%)	**90.8% (88.7%-92.7%)**
Air management	37.9% (35.9%-39.9%)	63.8% (61.7%-65.9%)	77.5% (75.6%-79.3%)	87.8% (86.3%-89.2%)	**92.5% (90.8%-93.9%)**	**93.2% (91.0%-94.9%)**	**95.0% (93.2%-96.3%)**
Proactive problem solving	22.2% (20.5%-23.9%)	39.5% (37.3%-41.7%)	57.9% (55.7%-60.1%)	72.4% (70.3%-74.3%)	82.0% (79.6%-84.2%)	86.7% (83.9%-89.1%)	88.1% (85.6%-90.2%)
Loop management	18.3% (16.7%-20.0%)	34.4% (32.3%-36.6%)	51.5% (49.2%-53.7%)	69.9% (67.8%-71.9%)	75.5% (72.9%-78.0%)	78.2% (74.9%-81.3%)	86.1% (83.5%-88.4%)
Patient comfort	40.0% (38.0%-42.0%)	62.0% (59.9%-64.2%)	77.0% (75.1%-78.8%)	85.5% (83.9%-87.0%)	**90.2% (88.4%-91.9%)**	**92.8% (90.7%-94.6%)**	**93.7% (91.8%-95.2%)**
Pace and progress	27.9% (26.1%-29.8%)	49.9% (47.7%-52.1%)	66.4% (64.3%-68.4%)	79.8% (78.0%-81.5%)	84.4% (82.2%-86.4%)	89.7% (87.2%-91.8%)	**91.2% (89.1%-93.1%)**

Visualisation	32.5% (30.6%-34.5%)	58.0% (55.8%-60.1%)	73.9% (71.9%-75.8%)	85.6% (84.0%-87.2%)	**90.9% (89.1%-92.5%)**	**93.6% (91.6%-95.3%)**	**96.1% (94.6%-97.3%)**
Recognition	42.0% (39.9%-44.2%)	62.0% (59.7%-64.2%)	78.8% (76.9%-80.6%)	89.1% (87.6%-90.5%)	**93.2% (91.6%-94.6%)**	**95.5% (93.7%-97.0%)**	**96.2% (94.6%-97.4%)**
Management	40.7% (38.4%-43.0%)	59.4% (57.0%-61.8%)	76.1% (74.0%-78.0%)	87.4% (85.8%-88.9%)	**91.4% (89.5%-93.0%)**	**94.7% (92.6%-96.3%)**	**93.7% (91.7%-95.4%)**
Complications	50.6% (46.5%-54.7%)	65.3% (61.2%-69.2%)	79.8% (76.6%-82.8%)	**90.5% (88.3%-92.4%)**	**96.3% (94.3%-97.7%)**	**97.6% (95.4%-98.9%)**	**95.9% (93.3%-97.7%)**
Report writing	53.6% (51.2%-56.0%)	74.0% (71.8%-76.2%)	86.1% (84.4%-87.7%)	**92.5% (91.2%-93.7%)**	**95.4% (94.0%-96.6%)**	**96.7% (95.1%-97.9%)**	**96.0% (94.4%-97.2%)**
Management plan	52.4% (50.1%-54.7%)	71.4% (69.2%-73.5%)	83.2% (81.4%-84.9%)	**92.0% (90.7%-93.2%)**	**95.0% (93.6%-96.2%)**	**94.7% (92.7%-96.2%)**	**95.6% (93.9%-96.9%)**
Communication and teamwork	61.6% (59.6%-63.6%)	77.5% (75.6%-79.3%)	85.9% (84.3%-87.4%)	**92.6% (91.3%-93.7%)**	**95.1% (93.7%-96.2%)**	**95.9% (94.2%-97.2%)**	**96.6% (95.2%-97.7%)**
Situation awareness	57.0% (54.9%-59.0%)	74.5% (72.5%-76.4%)	83.4% (81.7%-85.0%)	**91.6% (90.3%-92.8%)**	**94.8% (93.4%-96.0%)**	**95.8% (94.0%-97.1%)**	**94.9% (93.2%-96.3%)**
Leadership	52.2% (50.0%-54.3%)	68.2% (66.1%-70.3%)	80.6% (78.8%-82.3%)	89.1% (87.6%-90.4%)	**92.6% (91.0%-94.1%)**	**92.1% (89.9%-94.0%)**	**94.7% (93.0%-96.1%)**
Judgement and decision making	47.7% (45.6%-49.8%)	64.3% (62.1%-66.4%)	78.2% (76.3%-80.0%)	88.8% (87.3%-90.1%)	**91.4% (89.7%-93.0%)**	**93.4% (91.3%-95.1%)**	**94.8% (93.1%-96.2%)**
All items	50.7% (49.4%-52.1%)	68.7% (67.3%-70.0%)	80.1% (79.0%-81.3%)	89.1% (88.2%-89.9%)	**92.2% (91.2%-93.2%)**	**94.0% (92.9%-95.0%)**	**95.1% (94.2%-96.1%)**

Table 4: Competency rates as measured by colonoscopy DOPS. *The percentage of DOPS scoring competent for each assessed item (with 95% confidence intervals) is presented for each stratum of lifetime colonoscopy count. Highlighted values denote those which have reached the competency threshold of 90%.*

Factor	*N* (%)	Multivariable Odds Ratio	95% Confidence Interval	p-value
Trainee Specialty				
Gastroenterology	4469 (65.1%)	*REF*		
GI Surgeon	2399 (34.9%)	1.09	0.73-1.65	0.665
Trainee Seniority				0.220
Junior	2933 (42.7%)	*REF*		
Senior	3589 (52.3%)	1.26	0.97-1.64	0.091
Clinical Research Fellow	346 (5.0%)	1.25	0.77-2.03	0.372
Lifetime Colonoscopy Count				**<0.001**
<50	1256 (18.3%)	*REF*		
50-99	1267 (18.4%)	**2.97**	**2.18-4.03**	**<0.001**
100-149	1291 (18.8%)	**5.74**	**3.99-8.26**	**<0.001**
150-199	1358 (19.8%)	**11.5**	**7.67-17.1**	**<0.001**
200-249	720 (10.5%)	**16.7**	**10.3-27.0**	**<0.001**
250-299	440 (6.4%)	**25.9**	**14.1-47.2**	**<0.001**
300+	536 (7.8%)	**45.9**	**18.9-111.6**	**<0.001**
Assessor Role				**0.002**
Gastroenterologist	3911 (56.9%)	*REF*		
GI Surgeon	1597 (23.3%)	**1.51**	**1.08-2.10**	**0.001**
Non-medical endoscopist	1324 (19.3%)	**1.66**	**1.24-2.23**	**0.015**
Case Difficulty				**<0.001**
Easy	1582 (23.0%)	*REF*		
Moderate	4344 (63.2%)	**0.63**	**0.54-0.74**	**<0.001**
Complicated	942 (13.7%)	**0.42**	**0.34-0.52**	**<0.001**
Unassisted Caecal Intubation Rate (over 30 previous procedures)				**<0.001**
<70%	2467 (35.9%)	*REF*		
70-89%	2503 (36.4%)	**1.32**	**1.11-1.56**	**0.002**
90%+	1898 (27.6%)	**1.82**	**1.48-2.23**	**<0.001**
Lifetime Colonoscopy DOPS count				**0.002**
<5	2168 (31.6%)	*REF*		
5-9	1953 (28.4%)	**1.51**	**1.21-1.88**	**<0.001**
10-14	1222 (17.8%)	**1.55**	**1.14-2.10**	**0.005**
15-19	1525 (22.2%)	**1.85**	**1.26-2.72**	**0.002**
Gastroscopy Certification				
No	3104 (45.2%)	*REF*		
Yes	3764 (54.8%)	0.805	0.60-1.09	0.161
Lifetime Flexible Sigmoidoscopy Count				0.681
<30	2276 (33.1%)	*REF*		
30-39	2685 (39.1%)	1.05	0.82-1.35	0.680
80+	1907 (27.8%)	0.97	0.74-1.27	0.812

Table 5: Multivariable analysis of in-training factors associated with competence in colonoscopy DOPS (overall DOPS rating achieving the competent outcome), based on 6868 DOPS performed by gastroenterology and GI surgical trainees. *1st to 3rd year specialty trainees (ST3-ST5) were classified as junior seniority and ST6+ as senior.*

Generalised estimating equations were used to account for the non-independence of repeat DOPS by the same trainee. Bold p-values are significant at p<0.05.